

Мачульский Владислав Владимирович
Кубанский государственный университет

Сравнительный анализ алгоритмов YOLOv8 и YOLOv9 для задачи обнаружения объектов

Аннотация. В данной статье представлен сравнительный анализ моделей YOLOv8 и YOLOv9, предназначенных для автоматического обнаружения объектов на изображениях. Обе модели относятся к семейству YOLO, широко применяемому в задачах компьютерного зрения, однако различаются по ряду характеристик, влияющих на их практическую эффективность. В работе рассматриваются ключевые различия в скорости работы и точности распознавания, а также особенности применения каждой модели в зависимости от требований конкретных задач. YOLOv8 выделяется высокой производительностью и подходит для систем реального времени, таких как видеонаблюдение и автономные устройства. YOLOv9 ориентирована на повышение точности распознавания, что делает её предпочтительной в областях, где критически важна надёжность, например в медицине или промышленной диагностике. Представленный анализ может быть полезен при выборе подходящей модели для различных сфер применения в области компьютерного зрения.

Ключевые слова: искусственный интеллект, компьютерное зрение, обнаружение объектов, классификация объектов, алгоритмы детекции объектов, YOLO

Machulskii Vladislav Vladimirovich
Kuban State University

Comparative analysis of YOLOv8 and YOLOv9 algorithms for object detection tasks

Annotation. This article presents a comparative analysis of the YOLOv8 and YOLOv9 models designed for automatic object detection in images. Both models belong to the YOLO family, which is widely used in computer vision tasks, but differ in a number of characteristics that affect their practical efficiency. The article considers key differences in the speed and accuracy of recognition, as well as the features of using each model depending on the requirements of specific tasks. YOLOv8 stands out for its high performance and is suitable for real-time systems, such as video surveillance and autonomous devices. YOLOv9 is focused on improving recognition accuracy, which makes it preferable in areas where reliability is critical, such as medicine or industrial diagnostics. The presented analysis can be useful when choosing a suitable model for various computer vision applications.

Key words: artificial intelligence, computer vision, object detection, object classification, object detection algorithms, YOLO

YOLOv9 предлагает обновлённый подход к решению ключевых задач в области обнаружения объектов с применением глубоких нейронных сетей. Основное внимание уделяется вопросам потери информации и оптимизации архитектуры модели. В рамках этой концепции реализованы четыре ключевых элемента: Принцип информационного узкого места (The Information Bottleneck Principle), Обратимые функции (Reversible Functions), Программируемая Информация о Градиенте (Programmable Gradient Information (PGI)) и Обобщенная эффективная сеть агрегирования уровней (Generalized Efficient Layer Aggregate Network (GELAN)).

Принцип информационного узкого места отражает, как при прохождении данных через уровни нейросети часть информации теряется. Это описывается уравнением информационного узкого места.

$$I(X, X) \geq I(X, f_{\theta}(X)) \geq I(X, g_{\phi}(f_{\theta}(X))) \quad (1)$$

где

I – взаимная информация;

X – данные;

f_{θ} , g_{ϕ} – слои, через которые проходят данные.

Информационные потери могут привести к ухудшению качества градиентов и, как следствие, к проблемам с обучением модели. Для борьбы с этим YOLOv9 внедряет механизм PGI, который помогает сохранить важные характеристики данных на всех уровнях сети, обеспечивая стабильное обучение и более высокую точность.

Обратимость функций — ещё один важный элемент архитектуры. Обратимая функция позволяет восстановить исходные данные без потерь:

$$X = v_{\zeta}(r_{\psi}(X)) \quad (2)$$

где

X – данные;

ψ , ζ – параметры обратимой и ее обратной функции соответственно.

Это свойство способствует сохранению информации на всех этапах обработки, что особенно важно при построении глубоких сетей. YOLOv9 применяет обратимые функции, чтобы минимизировать утраты данных при переходе между слоями, тем самым повышая надёжность модели при обучении и инференсе.

Программируемая градиентная информация (PGI) представляет собой метод, в котором основная ветвь модели используется для вывода, а вспомогательная — для надёжного расчёта градиентов. Это решение обеспечивает точное обновление параметров модели, даже в лёгких и неглубоких архитектурах, не увеличивая вычислительные затраты при инференсе.

В связи с внедрением PGI возникает необходимость в архитектуре, способной эффективно поддерживать этот подход. Эту роль выполняет GELAN — обобщенная эффективная сеть агрегирования уровней, обеспечивающая гибкое и масштабируемое объединение вычислительных блоков. GELAN усиливает способность сети обрабатывать сложные паттерны и адаптироваться к различным задачам без ущерба для производительности.

Таким образом, сочетание PGI и GELAN в архитектуре YOLOv9 позволяет достигать высокого качества обнаружения объектов, сохраняя при этом вычислительную эффективность и универсальность модели.

И YOLOv8, и YOLOv9 наследуют фундаментальные принципы YOLO, но значительно различаются в своих конкретных архитектурных реализациях.

«Позвоночник» (Извлечение информативных свойств) сети формирует основу обеих моделей, на которую возложена фундаментальная ответственность за извлечение полных и отличительных признаков из входного изображения. YOLOv8 использует проверенную CSPDarknet32, включающую межэтапные частичные соединения (Cross-Stage Partial connections (CSP)), чтобы улучшить распространение градиента (gradient propagation) и сократить вычислительные требования. В свою очередь, YOLOv9 представляет YOLOv9-

CSPDarknet53, которая оптимизирует представление признаков для последующих задач обнаружения объектов.

«Шея» (Слияние признаков в разном масштабе). Слияние признаков разного масштаба необходимо для точного обнаружения объектов, особенно при работе с объектами разного масштаба. YOLOv8 развивает успех своих предшественников, используя Path Aggregation Network (PANet). PANet эффективно объединяет признаки с разных слоев, способствуя эффективному потоку семантической информации низкого и высокого уровня. YOLOv9 делает шаг вперед в инновациях, интегрируя Обобщенную эффективную сеть агрегирования уровней (GELAN). GELAN предоставляет большую гибкость за счет динамического выбора и агрегирования каналов, повышая способность модели изучать контекстуально релевантные признаки.

«Голова» (Прогнозирование объектов). Наконец, обе модели отвечают за генерацию окончательных координат ограничивающих рамок (bounding box) и вероятностей классов. Несмотря на структурную схожесть, YOLOv9 включает дополнительные «Фокусный» слой перед прогнозированием. Этот фокусный слой служит для улучшения масштабирования объектов и сохранения мелкозернистой информации, способствуя повышению точности обнаружения модели.

Различные механизмы объединения признаков и тонкие различия в «голове» прогнозирования способствуют расхождению в производительности обнаружения объектов между YOLOv8 и YOLOv9, особенно в точности и скорости.

Одно из самых существенных различий связано с точностью. YOLOv9 демонстрирует заметные улучшения в mean Average Precision (mAP) по сравнению с YOLOv8, обученной на известном датасете MS COCO. Это объясняется следующими факторами:

- Динамический отбор признаков GELAN позволяет модели сосредоточиться на информативных признаках, имеющих решающую роль для точного обнаружения объектов, потенциально уменьшая потерю информации по сравнению с YOLOv8, использующей подход статического слияния;

- Сохранение информации фокусного слоя служит для решения проблемы потери информации путем масштабирования признаков и сохранения деталей низкого уровня, необходимых для точной локализации ограничительных рамок (bounding box).

Хотя YOLOv9 отличается повышенной точностью, это стоит ей некоторого увеличения времени инференса по сравнению с YOLOv8. В основном это связано с введением динамических вычислений GELAN и дополнительных операций Фокусного слоя.

Итак, YOLOv8 отличается своей исключительной скоростью. Её эффективная архитектура позволяет сократить время инференса, что делает её идеальной для приложений реального времени, в которых скорость имеет решающее значение, таких как автономные транспортные средства, видеонаблюдение и обнаружение объектов с помощью БПЛА.

По сравнению с YOLOv9, YOLOv8 демонстрирует немного более низкую точность mAP на эталонных датасетах. Следовательно, YOLOv9 более предпочтительна для приложений, где точная идентификация объектов имеет принципиальное значение, таких как медицина, распознавание лиц и обнаружение дефектов на производстве.

Список источников

1. Computer Vision Algorithms : [сайт]. – 2024. – URL: <https://www.geeksforgeeks.org/computer-vision/computer-vision-algorithms/> (дата обращения: 25.03.2025).

2. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. – 2-е изд. – Москва : Издательство Юрайт, 2025. – 89 с. – ISBN 978-5-534-20732-3.

3. Шапиро, Л. Компьютерное зрение : учебник / Л. Шапиро, Д. Стокман. – 5-е изд. – Москва : Лаборатория знаний, 2024. – 762 с. – ISBN 978-5-93208-725-1.
4. Ultralytics YOLO Docs : [сайт]. – 2025. – URL: <https://docs.ultralytics.com/> (дата обращения: 27.03.2025).

Сведения об авторе

Мачульский Владислав Владимирович, бакалавр, «Информационные системы и технологии», ФГБОУ ВО «Кубанский государственный университет», г. Краснодар, Россия

Научный руководитель

Куликова Наталья Николаевна, кандидат биологических наук, доцент кафедры теоретической физики и компьютерных технологий, ФГБОУ ВО «Кубанский государственный университет», г. Краснодар, Россия

Information about the author

Machulsky Vladislav Vladimirovich, Bachelor's degree, "Information Systems and Technologies", Kuban State University

Scientific supervisor

Kulikova Natalia Nikolaevna, Candidate of Biological Sciences, Associate Professor of the Department of Theoretical Physics and Computer Technology, Kuban State University, Krasnodar, Russia